

ETSI TS 102 114 V1.1.1 (2002-08)

Technical Specification

DTS Coherent Acoustics; Core and Extensions

European Broadcasting Union



Union Européenne de Radio-Télévision

EBU·UER



Reference

DTS/JTC-DTS

Keywords

acoustic, audio, CODEC, coding, digital

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Individual copies of the present document can be downloaded from:

<http://www.etsi.org>

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at

<http://portal.etsi.org/tb/status/status.asp>

If you find errors in the present document, send your comment to:

editor@etsi.fr

Copyright Notification

No part may be reproduced except as authorized by written permission.
The copyright and the foregoing restriction extend to reproduction in all media.

© European Telecommunications Standards Institute 2002.

© European Broadcasting Union 2002.

All rights reserved.

DECTTM, **PLUGTESTS**TM and **UMTS**TM are Trade Marks of ETSI registered for the benefit of its Members.
TIPHONTM and the **TIPHON logo** are Trade Marks currently being registered by ETSI for the benefit of its Members.
3GPPTM is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners.

Contents

Intellectual Property Rights	4
Foreword.....	6
1 Scope	7
2 References	7
3 Definitions and abbreviations.....	7
3.1 Definitions	7
3.2 Abbreviations	7
4 Summary	8
5 Core Audio	8
5.1 Frame structure and decoding procedure.....	9
5.2 Error classification	10
5.3 Synchronization.....	11
5.4 Frame header	11
5.4.1 Bit stream header	11
6 Extension to more than 5.1 channels (XCh).....	19
6.1 Synchronization.....	19
6.2 Frame header	19
7 Extension to sampling frequencies of up to 96 kHz and/or higher resolution (X96k)	20
7.1 DTS Core+96 kHz-Extension encoder	21
7.2 DTS Core+96 kHz Extension decoder	22
7.3 Synchronization.....	22
7.4 X96k frame header	23
Annex A (informative): Bibliography.....	25
History	26

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: "*Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards*", which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (<http://webapp.etsi.org/IPR/home.asp>).

The attention of ETSI has been drawn to the Intellectual Property Rights (IPRs) listed below which are, or may be, or may become, Essential to the present document. The IPR owner (Digital Theatre Systems, Inc.) has undertaken to grant irrevocable licences, on fair, reasonable and non-discriminatory terms and conditions under these IPRs pursuant to the ETSI IPR Policy. The licensing undertaking has been made subject to the condition that those who seek licenses agree to reciprocate. Further details pertaining to these IPRs can be obtained directly from the IPR owner.

The present IPR information has been submitted to ETSI and pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

IPRs:

U.S. Patent No. 5,451,942 "Method and Apparatus for Multiplexed Encoding of Digital Audio Information Onto a Digital Audio Storage Medium"

- National patents and patent applications derived from PCT Application No. PCT/US95/00959, ruled patentable by International Preliminary Examination:
- Argentina Patent No. AR255019V1
- Australia Patent No. 680341
- Canada Patent No. 2,180,002
- Japan Patent No. 3187839
- Mexico Patent No. 184848
- South Africa Patent No. 95/0548
- Spain Patent No. 2115513
- Switzerland Patent No. 691 113

Additional applications:

- Brazil Application No. PI 9506695-0
- Chile Application No. 180.95
- China (PRC) Application No. 95191502.9
- European Patent Application No. 95908115.9 (France, Germany, Great Britain & Italy)
- Hong Kong Application No. 98114034.2
- India Application No. 1 686/Del/94
- Indonesia Application No. P-950164
- Korea Application No. 96-704141
- Malaysia Application No. PI9502624
- Philippines Application No. 51452

- Venezuela Application No. 0173-95

U.S. Patent No. 5,956,674 "A Multi-Channel Predictive Subband Audio Coder Using Psychoacoustic Adaptive Bit Allocation in Frequency, Time and Over the Multiple Channels."

U.S. Patent No. 5,974,380 "Multi-Channel Audio Decoder", a Divisional Application U.S. Patent No. 5,956,674.

U.S. Patent No. 5,978,762 "Digitally Encoded Machine Readable Storage Media Using Adaptive Bit Allocation in Frequency, Time and Over Multiple Channels", a Divisional Application U.S. Patent No. 5,956,674.

U.S. Patent Application No. 09/186,234 "Multi-Channel Audio Encoder", a Divisional Application U.S. Patent No. 5,956,674.

National patents and patent applications derived from PCT Application No. PCT/US96/18764:

- Australia Patent No. 705194
- Canada Patent No. 2,331,611
- Eurasia Patent No. 001087 (All Countries Designated)
- Korea Patent No. 0277819
- Taiwan Patent No. 92765

Additional applications:

- Brazil Application No. PI 9611852-0
- Canada Application No. 2,238,026 (division of Canada Patent No. 2,331,611)
- China Application No. 96199832.6
- European Patent Application No. 96941446.5 (All Countries Designated)
- Hong Kong Application No. 99100515.8
- India Patent Application No. 2592/DEL/96
- Japan Application No. 521314/97
- Mexico Patent Application No. 984320
- Poland Patent Application No. P-327 082, P-346 687 & P-346 688

U.S. Patent No.6,226,616 B1 "Sound Quality of Established Low Bit-Rate Audio Coding Systems without Loss of Decoder Compatibility".

National patents and patent applications derived from PCT Application No. PCT/US00/16681:

- China Application No. 00809269.9
- European Patent Application No. 00942890.5
- Hong Kong Application
- Japan Application No.
- Korea Application No. 10-2001-7016475
- Malaysia Application No. PI 20015600
- Singapore Application No. 200107899-9
- Taiwan Application No. 90131969

U.S. patent application Serial No. 09/568,355 "Discrete Multichannel Audio with a Backward Compatible Mix".

PCT application No. PCT/US01/14878, entitled "Discrete Multichannel Audio with a Backward Compatible Mix".

Any national applications derived from PCT application No. PCT/US01/14878, "Discrete Multichannel Audio with a Backward Compatible Mix."

Foreword

This Technical Specification (TS) has been produced by Joint Technical Committee (JTC) of the European Broadcasting Union (EBU), Comité Européen de Normalization ELECtrotechnique (CENELEC) and the European Telecommunications Standards Institute (ETSI).

NOTE: The EBU/ETSI JTC Broadcast was established in 1990 to co-ordinate the drafting of standards in the specific field of broadcasting and related fields. Since 1995 the JTC Broadcast became a tripartite body by including in the Memorandum of Understanding also CENELEC, which is responsible for the standardization of radio and television receivers. The EBU is a professional association of broadcasting organizations whose work includes the co-ordination of its members' activities in the technical, legal, programme-making and programme-exchange domains. The EBU has active members in about 60 countries in the European broadcasting area; its headquarters is in Geneva.

European Broadcasting Union
CH-1218 GRAND SACCONEX (Geneva)
Switzerland
Tel: +41 22 717 21 11
Fax: +41 22 717 24 81

1 Scope

The present document describes the key components of the DTS Coherent Acoustics technology. The document also includes the lists of all frame header parameters in the DTS core and extension (XCh and X96k) streams. The information about the remaining parameters of the DTS bit streams is further described in U.S. and other National patents which are listed in the Intellectual Property Rights clause of the present document, in connection with the intellectual property rights (IPRs) of DTS. These patents are published and are publicly available.

2 References

Void.

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

DTS Core Audio Stream: carries the coding parameters of up to 5.1 channels of the original LPCM audio at up to 24 bits per sample with the sampling frequency of up to 48 kHz

DTS Extended Audio Stream: delivers possible extended frequency bands of the primary audio channels as well as all frequency components of channels beyond 5.1.

NOTE: The extended audio stream must always have the accompanying core stream.

DTS XCh Stream: one of DTS extended streams that carries the coding parameters obtained from encoding of up to 2 additional channels of original LPCM audio at up to 24 bits per sample with the sampling frequency of up to 48 kHz

DTS X96k Stream: DTS extended audio stream that enables encoding of original LPCM audio at up to 24 bits per sample with the sampling frequency of up to 96 kHz

NOTE: The stream carries the coding parameters used for the representation of all remaining audio components that are present in the original LPCM audio and are not represented in the core audio stream

LPCM: Linear Pulse Code Modulated sequence of digital audio samples

QMF bank: specific filtering structure that provides the means of translating the time domain signal into the multiple sub-band domain signals

Vector Quantization: term for the joint quantization of a block of signal samples or a block of signal parameters

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

DTS	Digital Theatre Systems
LFE	Low Frequency Effect Channel
LPCM	Linear Pulse Code Modulation
QMF	Quadrature Mirror Filter
VQ	Vector Quantization

4 Summary

DTS Coherent Acoustics is designed to deliver digital audio reproduction in the home at studio quality level in terms of fidelity and sound stage imagery. Specifically, it delivers up to eight discrete channels of multiplexed audio at sampling frequencies of 8 kHz to 192 kHz at bit rates of 32 kbit/s to 6 144 kbit/s. The encoding algorithm works at 24 bits per sample and can deliver compression rate of 3:1 up to 40:1.

Due to the popularity of the 5.1 channel sound tracks in the movie industry and in the emerging multichannel home audio market, DTS Coherent Acoustics is delivered in the form of a core audio (for the 5.1 channels) plus optional extended audio (for the rest of the DTS Coherent Acoustics). The 5.1 channel audio consists of up to five primary audio channels with frequencies lower than 24 kHz plus a possible low frequency effect (LFE) channel (the 0.1 channel). This implies that the frequency components higher than 24 kHz for the five primary audio channels and all frequency components of the remaining two channels are carried in the extended audio. This structure is illustrated in figure 4.1 and as follows:

- Core Audio:
 - Up to 5 primary audio channels (frequency components below 24 kHz).
 - Up to 1 low frequency effect (LFE) channel.
 - Optional information such as time stamps and user information.
- Extended Audio:
 - Up to 2 additional full bandwidth channels (frequency components below 24 kHz).
 - Frequency components above 24 kHz for the primary and extended audio channels.

Under this structure, a basic DTS decoder can decode 5.1 channel core audio bits only and does not need to know even the existence of extended audio bits in the bit stream. A sophisticated decoder, however, can first decode the 5.1 core audio bits and then proceed to decode the extended audio bits if they exist.

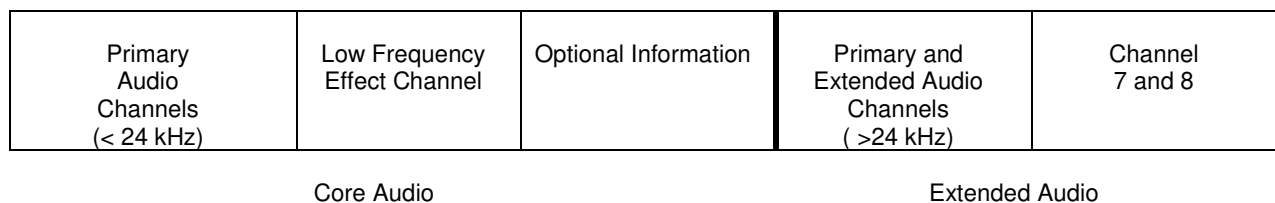


Figure 4.1: DTS Coherent Acoustics is optimized for 5.1 channel applications, but is extensible to deliver 8 channels with sampling frequency up to 192 kHz

5 Core Audio

DTS core encoder delivers 5.1 channel audio at 24 bits per sample with a sampling frequency of up to 48 kHz. As shown in figure 5.1, the audio samples of a primary channel are split and decimated by a 32-band QMF bank into 32 sub-bands. The samples of each sub-band goes through an adaptive prediction process to check if the resultant prediction gain is large enough to justify the overhead of transferring the coefficients of prediction filter. The prediction gain is obtained by comparing the variance of the prediction residual to that of the sub-band samples. If the prediction gain is big enough, the prediction residual is quantified using mid-tread scalar quantization and the prediction coefficients are vector-quantized (VQ). Otherwise, the sub-band samples themselves are quantized using mid-tread scalar quantization. In the case of low bit rate applications, the scalar quantization indexes of the residual or sub-band samples are further encoded using Huffman code. When the bit rate is low, vector quantization (VQ) may also be used to quantize samples of the high-frequency sub-bands for which the adaptive prediction is disabled. In very low bit rate applications, joint intensity coding and sum/difference coding may be employed to further improve audio quality. The optional LFE channel is compressed by: low-pass filtering, decimation and mid-tread scalar quantization.

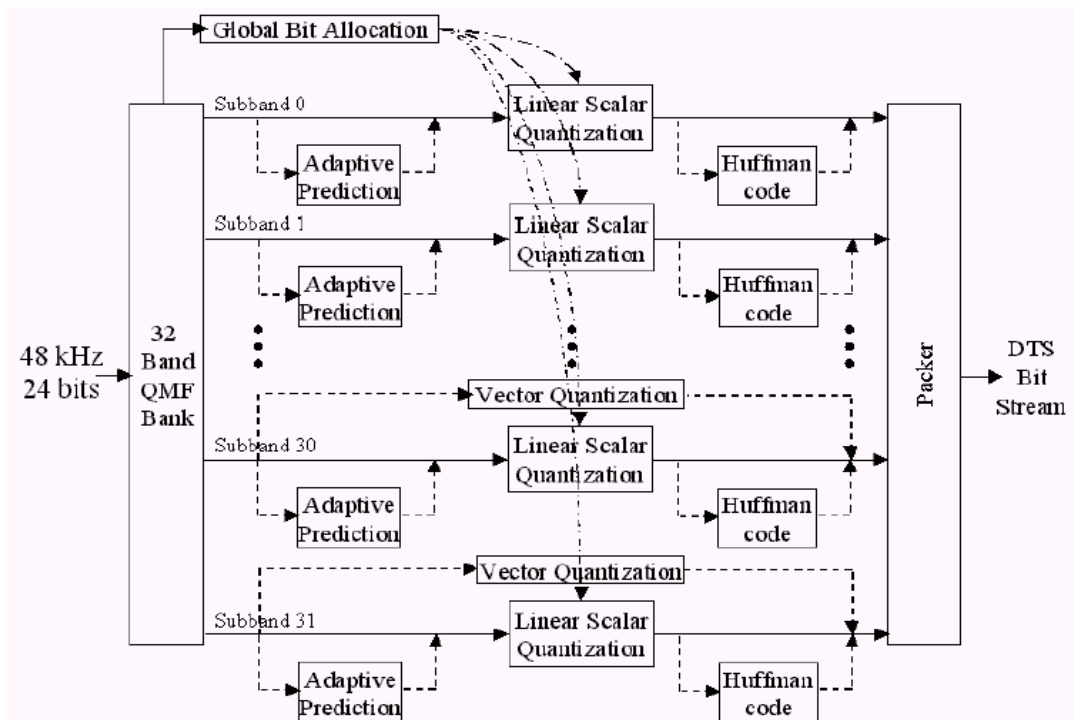


Figure 5.1: Compression of a primary audio channel. The dotted lines indicate optional operations and dash dot lines bit allocation control

5.1 Frame structure and decoding procedure

DTS bit stream is a sequence of synchronized frames, each consisting of the following fields (see figure 5.2):

- **Synchronization Word:** Synchronize the decoder to the bit stream.
- **Frame Header:** Carries information about frame construction, encoder configuration, audio data arrangement, and various operational features.
- **Sub-frames:** Carries core audio data for the 5.1 channels. Each frame may have up to 16 sub-frames.
- **Optional Information:** Carries auxiliary data such as time code, which is not intrinsic to the operation of the decoder but may be used for post processing routines.
- **Extended Audio:** Carries possible extended frequency bands of the primary audio channels as well as all frequency components of channels beyond 5.1.

Each sub-frame contains data for audio samples of the 5.1 channels covering a time duration of up to that of the sub-band analysis window and can be decoded entirely without reference to any other sub-frames. A sub-frame consists of the following fields (see figure 5.3):

- **Side Information:** Relays information about how to decode the 5.1 channel audio data. Information for joint intensity coding is also included here.
- **High Frequency VQ:** Some and a small number of high frequency sub-bands of the primary channels may be encoded using VQ. In this case, the samples of each of those sub-bands within the sub-frame are encoded as a single VQ address.
- **Low Frequency Effect Channel:** The decimated samples of the LFE channel are carried as 8-bit words.
- **Sub-sub-frames:** All sub-bands, except those high-frequency VQ encoded ones, are encoded here in up to 4 sub-sub-frames.

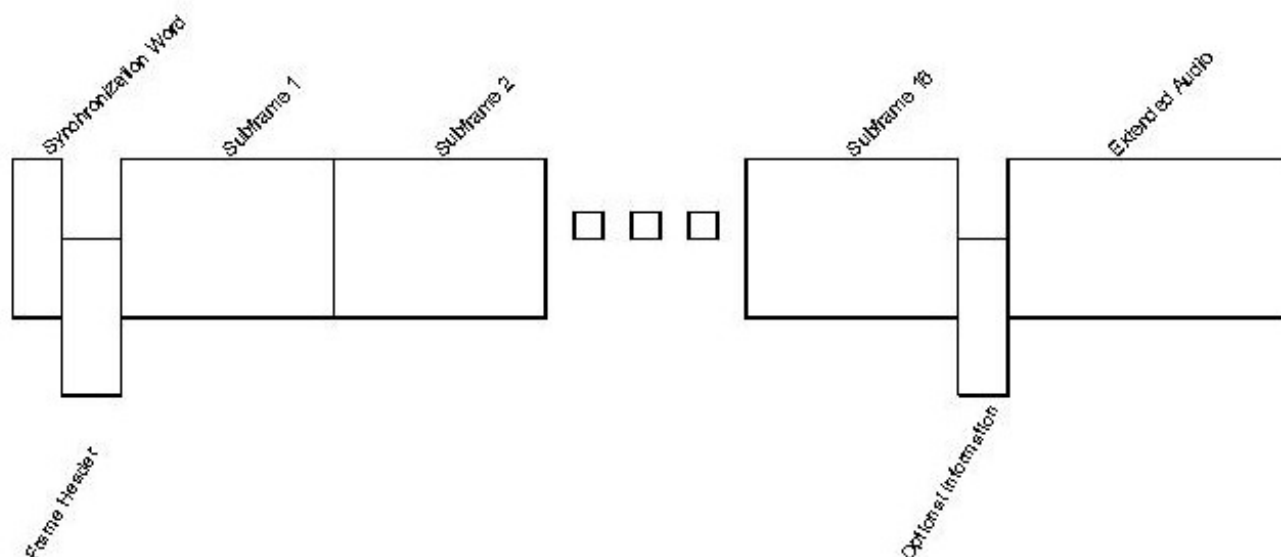


Figure 5.2: DTS frame structure

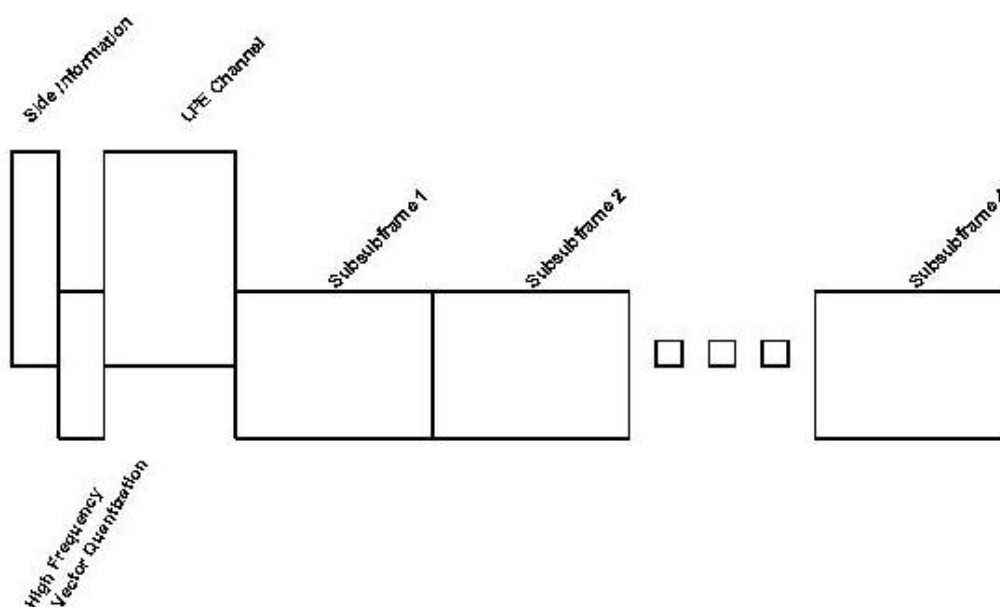


Figure 5.3: Sub-frame structure

5.2 Error classification

Each element in the bit stream carries either a piece of the audio data or the information to decode them. A corrupted bit stream element will cause an error in the decoder and its consequences depend on the information that element carries. In order to control decoded audio quality, the consequence of a corrupted element is categorized as

- V** Vital: The element is designed to change from frame to frame and its corruption is likely to lead to failure in the decoding process and instability in decoded PCM outputs.
- ACC** Corruption could cause failure. Since the element usually does not change from frame to frame, the error may be compensated for by a majority vote over consecutive frames.
- NV** Non-vital: corruption will degrade the quality of PCM outputs, but the degradation will be graceful.

5.3 Synchronization

DTS bit stream consists of a sequence of audio frames of equal size, each begins with a 32-bit synchronization word:

SYNC = 0x7ffe8001 **V** **32 bits**

So the first decoding step is to search the input bit stream for SYNC. In order to reduce the probability of false synchronization, 6 bits after SYNC in the bit stream may be further checked, since they usually do not change for normal frames (they do carry useful information about frame structure). These 6 bits should be 0x3f (the binary 111111) for normal frames and are called synchronization word extension. Concatenating them with SYNC gives an extended synchronization word (32 + 6 = 38 bits):

SYNC = 0x7ffe8001 + 0x3f for normal frame **V** **38 bits**

which reduces the probability of false synchronization to 10^{-7} . In addition, the fact that SYNC occurs at a fixed interval further reduces the probability of false synchronization to almost zero.

The above search procedure shall be carried out only when the decoder is out of synchronization with the bit stream. After synchronization is established, the decoder should only check if **SYNC = 0x7ffe8001** before it begins to decode a frame, because the 6 bits after SYNC may change for abnormal (termination) frames.

The SYNC word appears at the beginning of each DTS data frame in the stream. The length of the DTS data frame is fixed for the entire DTS stream and consequently the SYNC words occur at the fixed intervals within the stream. During the initial synchronization process the decoder shall calculate the distance between the two consecutive SYNC words. While in synchronization with the incoming DTS stream, the decoder shall only look for the SYNC word of a new data frame at the calculated distance from the SYNC word of previously decoded data frame. If the SYNC word is found at the specified distance the decoder shall proceed with the decoding of the new data frame and if not the "out-of-sync" state shall be pronounced.

When DTS bit stream is stored in 16-bit words such as on CD, SYNC will be stored as 0x7ffe and 0x8001. However, when DTS bit stream is viewed on an IBM PC platform, since the high byte and low byte are switched, SYNC will appear like 0xfe7f and x0180.

Note that, in order to make the harsh sound less unpleasant when DTS bit stream is mistakenly played back as PCM format, DTS now provides a 14-bit format that reduces the dynamic range from 16 to 14 bits. In this 14-bit format, DTS bit stream is stored only in the least significant 14 bits of a 16-bit word, the most significant 2 bits are not used, In case of this, SYNC is stored in three words: 0x1fff, 0xe800, and 0x07f.

5.4 Frame header

The frame header consists of a bit stream header and a primary audio coding header. The bit stream header provides information about the construction of the frame, the encoder configuration such as core source sampling frequency, and various optional operational features such as embedded dynamic range control. The primary audio coding header specifies the packing arrangement and coding formats used at the encoder to assemble the audio coding side information. Many elements in the headers are repeated for each separate audio channel.

5.4.1 Bit stream header

Frame Type **V** **FTYPE** **1 bit**

It indicates the type of current frame:

Table 5.1: Frame Type

FTYPE	Frame Type
1	Normal frame
0	Termination frame

Termination frames are used when it is necessary to accurately align the end of an audio sequence with a video frame end point. A termination block carries $n \times 32$ core audio samples where block length n is adjusted to just fall short of the video end point. Two termination frames may be transmitted sequentially to avoid transmitting one excessively small frame.

Deficit Sample Count **V** **SHORT** **5 bits**

It defines the number of core samples by which a termination frame falls SHORT of the normal length of a block. A block = 32 PCM core samples per channel, corresponding to the number of PCM core samples that are feed to the core filter bank to generate one sub-band sample for each sub-band. A normal frame consists of blocks of 32 PCM core samples, while a termination frame provides the flexibility of having a frame size precision finer than the 32 PCM core sample block. On completion of a termination frame, (SHORT+1) PCM core samples must be padded to the output buffers of each channel. The padded samples may be zeros or they may be copies of adjacent samples.

Table 5.2: Deficit Sample Count

SHORT	Valid Value or Range of SHORT
1	[0,30]
0	31 (indicating a normal frame).

CRC Present Flag **V** **CPF** **1 bit**

A flag that indicates if CRC (cyclic redundancy check) bits present in the bit stream.

Table 5.3: CRC Present Flag

CPF	CRC
1	Present
0	Not Present

Number of PCM Sample Blocks **V** **NBLKS** **7 bits**

It indicates that there are (NBLKS + 1) blocks (a block = 32 PCM core samples per channel, corresponding to the number of PCM samples that are fed to the core filter bank to generate one sub-band sample for each sub-band) in the current frame (see note). The actual core encoding window size is $32 \times (\text{NBLKS} + 1)$ PCM samples per channel. Valid range for NBLKS: 5 to 127. Invalid range for NBLKS: 0 to 4. For normal frames, this indicates a window size of either 2 048, 1 024, 512, or 256 samples per channel. For termination frames, NBLKS can take any value in its valid range.

NOTE: When frequency extension stream (X96k) is present, the PCM core samples represent the samples at the output of the decimator that precedes the core encoder. This k -times decimator translates the original PCM source samples with the sampling frequency of $F_{s_src} = k \times \text{SFREQ}$ to the core PCM samples ($F_{s_core} = \text{SFREQ}$) suitable for the encoding by the core encoder. The core encoder can handle sampling frequencies $\text{SFREQ} \leq 48 \text{ kHz}$ and consequently;

- $k = 2$ for $48 \text{ kHz} < F_{src} \leq 96 \text{ kHz}$ and
- $k = 4$ for $96 \text{ kHz} < F_{src} \leq 192 \text{ kHz}$

Primary Frame Byte Size **V** **FSIZE** **14 bits**

(FSIZE+1) is the total byte size of the current frame including primary audio data as well as any extension audio data. Valid range for FSIZE: 95 to 16 383. Invalid range for FSIZE: 0 to 94.

Audio Channel Arrangement **ACC** **AMODE** **6 bits**

Audio channel arrangement that describes the number of audio channels (CHS) and the audio playback arrangement (see table 5.4). Unspecified modes may be defined at a later date (user defined code) and the control data required to implement them, i.e. channel assignments, down mixing etc, can be uploaded from the player platform.

Table 5.4: Audio channel arrangement

AMODE	CHS	Arrangement
0b000000	1	A
0b000001	2	A + B (dual mono)
0b000010	2	L + R (stereo)
0b000011	2	(L + R) + (L - R) (sum - difference)
0b000100	2	LT + RT (left and right total)
0b000101	3	C + L + R
0b000110	3	L + R + S
0b000111	4	C + L + R + S
0b001000	4	L + R + SL + SR
0b001001	5	C + L + R + SL + SR
0b001010	6	CL + CR + L + R + SL + SR
0b001011	6	C + L + R + LR + RR + OV
0b001100	6	CF + CR + LF + RF + LR + RR
0b001101	7	CL + C + CR + L + R + SL + SR
0b001110	8	CL + CR + L + R + SL1 + SL2 + SR1 + SR2
0b001111	8	CL + C + CR + L + R + SL + S + SR
0b010000 - 0b111111		User defined

Legends: L = left, R = right, C = center, S = surround, F = front, R = rear, T = total, OV = overhead

Core Audio Sampling Frequency ACC SFREQ 4 bits

It specifies the sampling frequency of audio samples in the core encoder, based on table 5.5. When the source sampling frequency is beyond 48 kHz the audio is encoded in up to 3 separate frequency bands. The base-band audio, for example, 0 kHz to 16 kHz, 0 kHz to 22,05 kHz or 0 kHz to 24 kHz, is encoded and packed into the core audio data arrays. The SFREQ corresponds to the sampling frequency of the base-band audio. The audio above the base-band (the extended bands), for example, 16 kHz to 32kHz, 22,05 kHz to 44,1 kHz, 24 kHz to 48 kHz, is encoded and packed into the extended coding arrays which reside at the end of the core audio data arrays. If the decoder is unable to make use of the high sample rate data this information may be ignored and the base-band audio converted normally using a standard sampling rates (32 kHz, 44,1 kHz or 48 kHz). If the decoder is receiving data coded at sampling rates lower than that available from the system then interpolation (2× or 4×) will be required (see table 5.6).

Table 5.5: Core audio sampling frequencies

SFREQ	Core Audio Sampling Frequency
0b0000	Invalid
0b0001	8 kHz
0b0010	16 kHz
0b0011	32 kHz
0b0100	Invalid
0b0101	Invalid
0b0110	11,025 kHz
0b0111	22,05 kHz
0b1000	44,1 kHz
0b1001	Invalid
0b1010	Invalid
0b1011	12 kHz
0b1100	24 kHz
0b1101	48 kHz
0b1110	Invalid
0b1111	Invalid

Table 5.6: Sub-sampled audio decoding for standard sampling rates

Core Audio Sampling Frequency	Hardware Sampling Frequency	Required Filtering
8 kHz	32 kHz	4 × Interpolation
16 kHz	32 kHz	2 × Interpolation
32 kHz	32 kHz	none
11 kHz	44,1 kHz	4 × Interpolation
22,05 kHz	44,1 kHz	2 × Interpolation
44,1 kHz	44,1 kHz	none
12 kHz	48 kHz	4 × Interpolation
24 kHz	48 kHz	2 × Interpolation
48 kHz	48 kHz	none

Transmission Bit Rate **ACC** **RATE** **5 bits**

RATE specifies the targeted transmission data rate for the current frame of audio (see table 5.7). The open mode allows for bit rates not defined by the table. Variable and loss-less modes imply that the data rate changes from frame to frame.

Table 5.7: RATE parameter vs. targeted bit-rate

RATE	Targeted Bit Rate [kbit/s]
0b00000	32
0b00001	56
0b00010	64
0b00011	96
0b00100	112
0b00101	128
0b00110	192
0b00111	224
0b01000	256
0b01001	320
0b01010	384
0b01011	448
0b01100	512
0b01101	576
0b01110	640
0b01111	768
0b10000	960
0b10001	1 024
0b10010	1 152
0b10011	1 280
0b10100	1 344
0b10101	1 408
0b10110	1 411,2
0b10111	1 472
0b11000	1 536
0b11001	1 920
0b11010	2 048
0b11011	3 072
0b11100	3 840
0b11101	open
0b11110	Variable
0b11111	Loss-less

Due to the limitations of the transmission medium the actual bit rate may be slightly different from the targeted bit rate, as listed in table 5.8 for the two types of applications. The bit-rates that are not shown in the table 5.8 are not applicable on either of these two applications.

Table 5.8: Targeted and actual bit-rate for the CD and DVD-Video applications

RATE	Targeted Bit Rate [kbit/s]	Actual Bit Rate on DTS CDs [kbit/s]		Actual Bit Rate on DVD-Video Discs [kbit/s]
		14-bit format	16-bit format	
0b01111	768	N/A	N/A	754,50
0b10110	1 411,2	1 234,8	1 411,2	N/A
0b11000	1 536	N/A	N/A	1 509,75

Embedded Down Mix Enabled V MIX 1 bit

This indicates if embedded down mixing coefficients are included at the start of each sub-frame (see table 5.9). Down mixing to stereo may be implemented using these coefficients for the duration of the sub-frame.

Table 5.9: Status of embedded down mixing coefficients

MIX	Mix Parameters
0	not present
1	present

Embedded Dynamic Range Flag V DYNF 1 bit

DYNF indicates if embedded dynamic range coefficients are included at the start of each sub-frame. Dynamic range correction may be implemented on all channels using these coefficients for the duration of the sub-frame.

Table 5.10: Embedded Dynamic Range Flag

DYNF	Dynamic Range Coefficients
0	not present
1	present

Embedded Time Stamp Flag V TIMEF 1 bit

It indicates if embedded time stamps are included at the end of the core audio data.

Table 5.11: Embedded Time Stamp Flag

TIMEF	Time Stamps
0	not present
1	present

Auxiliary Data Flag V AUXF 1 bit

It indicates if auxiliary data bytes are appended at the end of the core audio data.

Table 5.12: Auxiliary Data Flag

AUXF	Auxiliary Data Bytes
0	not present
1	present

HDCD **NV** **HDCD** **1 bits**

The source material is mastered in HDCD format if HDCD = 1, and otherwise HDCD = 0.

Extension Audio Descriptor Flag **ACC** **EXT_AUDIO_ID** **3 bits**

This flag has meaning only if the EXT_AUDIO = 1 (see below) and then it indicates the type of data that has been placed in the extension stream(s).

Table 5.13: Extension Audio Descriptor Flag

EXT_AUDIO_ID	Type of Extension Data
0	Channel Extension (XCh)
1	Reserved
2	Frequency Extension (X96k)
3	XCh and X96k
4	Reserved
5	Reserved
6	Reserved
7	Reserved

Extended Coding Flag **ACC** **EXT_AUDIO** **1 bit**

It indicates if extended audio coding data are present after the core audio data. Extended audio data will include the data for the extended bands of the 5 normal primary channels as well as all bands of additional audio channels. To simplify the process of implementing a 5,1ch/48 kHz decoder, the extended coding data arrays are placed at the end of the core audio array.

Table 5.14: Extended Coding Flag

EXT_AUDIO	Extended Audio Data
0	not present
1	present

Audio Sync Word Insertion Flag **ACC** **ASPF** **1 bit**

It indicates how often the audio data check word DSYNC (0xFFFF Extension Audio Descriptor Flag) occurs in the data stream. DSYNC is used as a simple means of detecting the presence of bit errors in the bit stream and is used as the final data verification stage prior to transmitting the reconstructed PCM words to the DACs.

Table 5.15: Audio Sync Word Insertion Flag

ASPF	DSYNC Placed at End of Each
0	Sub-frame
1	Sub-sub-frame

Low Frequency Effects Flag **V** **LFF** **2 bits**

Indicates if the LFE channel is present and the choice of the interpolation factor to reconstruct the LFE channel (see table 5.16).

Table 5.16: Flag for LFE channel

LFF	LFE Channel	Interpolation Factor
0	not present	
1	Present	128
2	Present	64
3	Invalid	

Predictor History Flag Switch **V** **HFLAG** **1 bit**

If frames are to be used as possible entry points into the data stream or as audio sequence "start frames" the ADPCM predictor history may not be contiguous. Hence these frames can be coded without the previous frame predictor history, making audio ramp-up faster on entry. When generating ADPCM predictions for current frame, the decoder will use reconstruction history of the previous frame if HFLAG = 1. Otherwise, the history will be ignored.

Header CRC Check Bytes **V** **HCRC** **16 bits**

This 16-bit CRC check word checks if there are errors from beginning of the current frame up to this point. It is present only if CPF = 1.

Multirate Interpolator Switch **NV** **FILTS** **1 bit**

This flag indicates which set of 32-band interpolation FIR coefficients is to be used to reconstruct the sub-band audio (see table 5.17).

Table 5.17: Multirate interpolation filter bank switch

FILTS	32-band Interpolation Filter
0	Non-perfect Reconstruction
1	Perfect Reconstruction

Encoder Software Revision **ACC/NV** **VERNUM** **4 bits**

It indicates of the revision status of the encoder software (see table 5.18). In addition the VERNUM is used to indicate the presence of the dialog normalization parameters (see table 5.22).

Table 5.18: Encoder software revision

VERNUM	Encoder Software Revision
0 to 6	Future revision (compatible with the present document)
7	Current
8 to 15	Future revision (incompatible with the present document)

NOTE: If the decoder encounters the DTS stream with the VERNUM >7 and the decoder is not designed for that specific encoder software revision than it must mute its outputs.

Copy History **NV** **CHIST** **2 bits**

It indicates the copy history of the audio. Because of the copyright regulations, the exact definition of this field is deliberately omitted.

Source PCM Resolution **ACC/NV** **PCMR** **3 bits**

It indicates the quantization resolution of source PCM samples (see table 5.19). The left and right surrounding channels of the source material are mastered in DTS ES format if ES = 1, and otherwise if ES = 0.

Table 5.19: Quantization resolution of source PCM samples

PCMR	Source PCM Resolution	ES
0b000	16 bits	0
0b001	16 bits	1
0b010	20 bits	0
0b011	20 bits	1
0b110	24 bits	0
0b101	24 bits	1
Others	Invalid	invalid

Front Sum/Difference Flag **V** **SUMF** **1 bit**

Indicates if front left and right channels are sum-difference encoded prior to encoding (see table 5.20). If set to zero no decoding post processing is required at the decoder.

Table 5.20: Sum/difference decoding status of front left and right channels

SUMF	Front Sum/Difference Encoding
0	$L = L, R = R$
1	$L = L + R, R = L - R$

Surrounds Sum/Difference Flag **V** **SUMS** **1 bit**

Indicates if left and right surround channels are sum-difference encoded prior to encoding (see table 5.21). If set to zero no decoding post processing is required at the decoder.

Table 5.21: Sum/difference decoding status of left and right surround channels

SUMS	Surround Sum/Difference Encoding
0	$Ls = Ls, Rs = Rs$
1	$Ls = Ls + Rs, Rs = Ls - Rs$

Dialog Normalization Parameter/Unspecified **V** **DIALNORM/UNSPEC** **4 bits**

For the values of VERNUM = 6 or 7 this 4-bit field is used to determine the dialog normalization parameter. For all other values of the VERNUM this field is a place holder that is not specified at this time.

The dialog normalization gain (DNG), in dB, is specified by the encoder operator and is used to directly scale the decoder outputs samples. In the DTS stream the information about the DNG value is transmitted by means of combined data in the VERNUM and DIALNORM fields (see table 5.22).

For all other values of the VERNUM (i.e. 0, 1, 2, 3, 4, 5, 8, 9, ...15) the UNSPEC 4-bit field should be extracted but ignored by the decoder. In addition, for these VERNUM values, the dialog normalization gain should be set to 0 i.e., DNG = 0 -> No Dialog Normalization.

Table 5.22: Dialog Normalization Parameter

Dialog Normalization Gain (DNG) Applied to the Decoder Outputs [dB]	VERNUM	DIALNORM
0	7	0b0000
-1	7	0b0001
-2	7	0b0010
-3	7	0b0011
-4	7	0b0100
-5	7	0b0101
-6	7	0b0110
-7	7	0b0111
-8	7	0b1000
-9	7	0b1001
-10	7	0b1010
-11	7	0b1011
-12	7	0b1100
-13	7	0b1101
-14	7	0b1110
-15	7	0b1111
-16	6	0b0000
-17	6	0b0001
-18	6	0b0010
-19	6	0b0011
-20	6	0b0100
-21	6	0b0101
-22	6	0b0110

Dialog Normalization Gain (DNG) Applied to the Decoder Outputs [dB]	VERNUM	DIALNORM
-23	6	0b0111
-24	6	0b1000
-25	6	0b1001
-26	6	0b1010
-27	6	0b1011
-28	6	0b1100
-29	6	0b1101
-30	6	0b1110
-31	6	0b1111

6 Extension to more than 5.1 channels (XCh)

When the need arises to encode more than 5.1 channels, the extended channels are compressed using exactly the same technology as the core audio channels. The audio data representing these extension channels are appended to the end of the DTS stream audio. These extension audio data are automatically ignored by the first generation DTS decoders but can be decoded by the second generation DTS decoders. The decoding process flows as follows.

6.1 Synchronization

Channel Extension Sync Word V **XChSYNC** **32 bits**

The synchronization word XChSYNC = 0x5a5a5a5a for the channel extension audio comes after all other extension streams i.e., in case of multiple extension streams the XCh stream is always the last. For 16 bit streams, XChSYNC is aligned to 32-bit word boundary. For 14 bit streams, it is aligned to both 32 bit and 28 bit word boundaries, meaning that, the sync word appears as 0x1696e5a5 in the 28 bit stream and as 0x5a5a5a5a after this stream is packed into a 32 bit stream.

Since the pseudo sync word might appear in the bit stream, it is MANDATORY to check the distance between this sync and the end of the encoded bit stream. This distance in bytes should be equal to XChFSIZE+1. The parameter XChFSIZE is described below.

NOTE: For compatibility reasons with legacy bit streams the estimated distance in bytes is checked against both the XChFSIZE+1 as well as the XChFSIZE. The XCh synchronization is pronounced only if the distance matches either of these two values.

6.2 Frame header

Primary Frame Byte Size V **XChFSIZE** **10 bits**

(XChFSIZE+1) is the distance in bytes from current extension sync word to the end of the current audio frame. Valid range for XChFSIZE: 95 to 1 023. Invalid range for XChFSIZE: 0 to 94.

Extension Channel Arrangement ACC **AMODE** **4 bits**

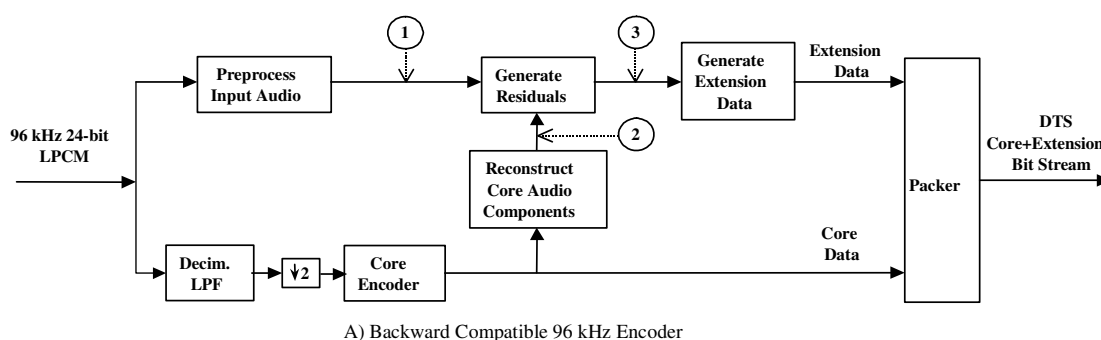
Audio channel arrangement that describes the number of audio channels (CHS) and the audio playback arrangement. It is set to represent the number of extension channels for now. More detail will be added in the future.

7 Extension to sampling frequencies of up to 96 kHz and/or higher resolution (X96k)

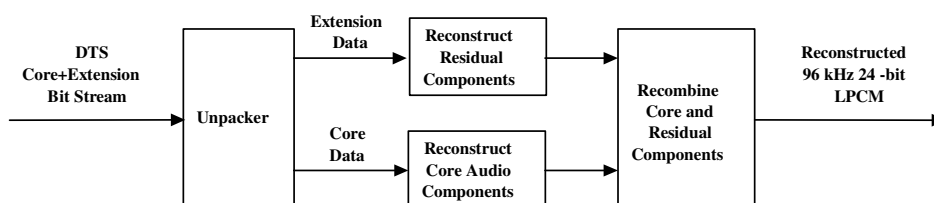
The generalized concept of core + 96 kHz-extension coding is illustrated in figure 7.1. To encode 96 kHz LPCM the input audio stream is fed to a 96 kHz to 48 kHz down sampler and the resulting 48 kHz signal is encoded using standard core encoder as in figure 7.1A). Referring to figure 7.1A):

- In the "Preprocess Input Audio" block the original 96 kHz/24-bit LPCM audio is first delayed and next passed through the extension 64-band analysis filter bank. Signal "1" in this case consists of the extension sub-band samples @ 96 kHz/64.
- The core data consists of the core audio codes in 32 sub-bands and the side information. In the "Reconstruct Core Audio Components" block the core audio codes are inverse quantized to produce the reconstructed core sub-band samples @ 48 kHz/32. These sub-band samples correspond to signal "2".
- In the "Generate Residuals" block the reconstructed core sub-band samples are subtracted from the extension sub-band samples in the lower 32 sub-bands. The extension sub-band samples in the upper 32 bands remain unaltered. These residual sub-band samples in the 64 bands correspond to signal "3".
- The ("Generate Extension Data" block processes the residual sub-band samples and generates the extension data that, along with the core data, is assembled in a packer to produce a core+extension bit stream.

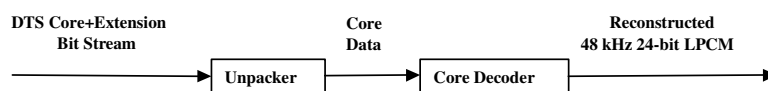
In the 96 kHz decoder, figure 7.1B), the unpacker first separates the core+extension stream into the core and extension data. The core sub-band decoder, in the "Reconstruct Core Audio Components" block, processes the core data and produces the reconstructed core sub-band samples (same as signal "2" generated in the encoder). Next in the "Reconstruct Residual Components" block, the extension sub-band decoder uses the extension data to generate the reconstructed residual sub-band samples in the 64 bands. In the "Recombine Core and Residual Components" block the core sub-band samples are added to the lower 32 bands of residual sub-band samples to produce the extension sub-band samples in the 64 bands. In the same block the synthesis 64-band filter bank processes the extension sub-band samples and generates the 96 kHz 24-bit LPCM audio. The combining of reconstructed residuals and core signals on the decoder side, figure 7.1B), is also done in sub-band domain.



A) Backward Compatible 96 kHz Encoder



B) 96 kHz Decoder



C) 48 kHz (Legacy) Decoder

Figure 7.1: The concept of Core+Extension coding methodology

When a 48 kHz-only (legacy) decoder is fed the core + extension bit stream, figure 7.1C), the extension data fields are ignored and only the core data is decoded. This results in 48 kHz core LPCM audio output.

7.1 DTS Core+96 kHz-Extension encoder

The block diagram in figure 7.2 shows the main components of the encoding algorithm. The input digital audio signal with a sampling frequency up to 96 kHz and a word length up to 24 bits is processed in the core branch and extension branch. In the core branch input audio is low-pass filtered to reduce its bandwidth to below 24 kHz, and then decimated by a factor of two, resulting in a 48 kHz sampled audio signal. The purpose of this LPF decimation is to remove signal components that cannot be represented by the core algorithm. The down sampled audio signal is processed in a 32-band analysis cosine modulated filter bank that produces the core sub-band samples. The core bit allocation routine based on the energy contained in each of the sub-bands and configuration of the core encoder determines the desired quantization scheme for each of the sub-bands. The core sub-band encoder performs quantization and encoding after which the audio codes and side information are delivered to the packer. The packer assembles this data into a core bit stream.

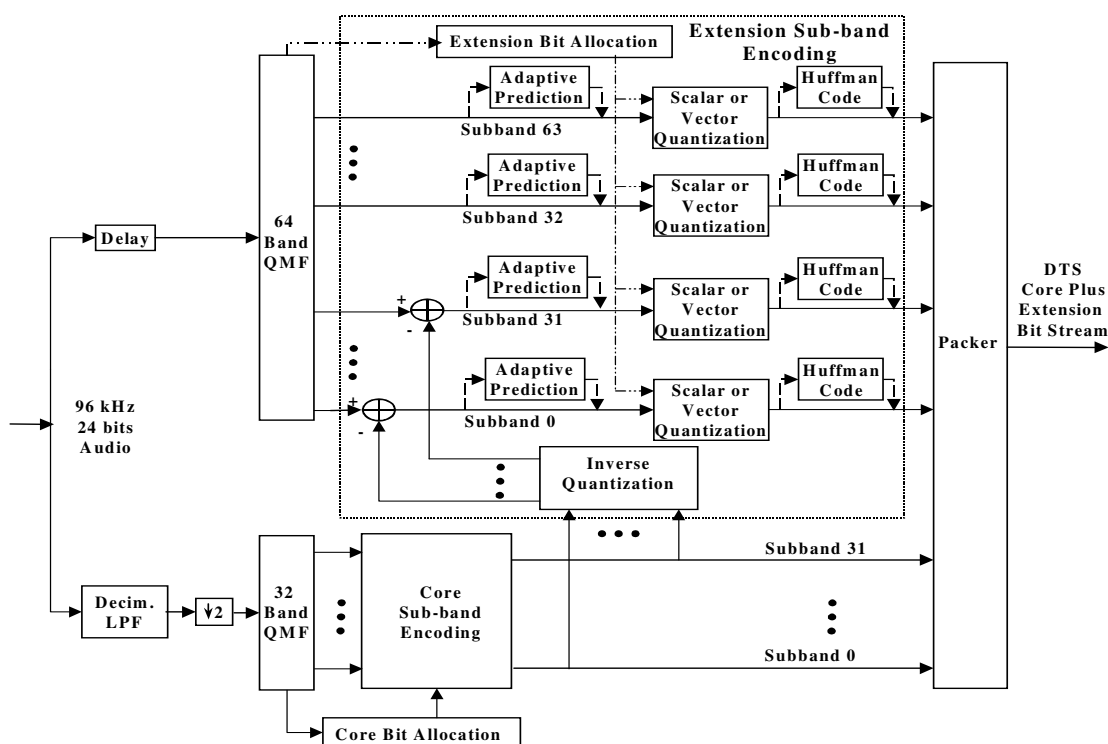


Figure 7.2: The block diagram of DTS Core+Extension encoder

In the extension branch the delayed version of input audio is processed in a 64-band analysis cosine modulated filter bank that produces the extension sub-band samples. Inverse quantization of the core audio codes produces the reconstructed core sub-band samples. Subtracting these samples from the extension sub-band samples in the lower 32 bands generates the residual sub-band samples. The residual signals in the upper 32 sub-bands are unaltered extension sub-band samples in corresponding bands. The delay of input audio is such that reconstructed core sub-band samples and extension sub-band samples in the lower 32 bands are time-aligned before the residual signals are produced i.e.,

$$\text{Delay} = \text{Delay}_{\text{DecimationLPF}} + \text{Delay}_{\text{CoreQMF}} - \text{Delay}_{\text{ExtensionQMF}}$$

The extension bit allocation routine based on the energy of residuals in each of the sub-bands and configuration of the extension encoder determines the desired quantization scheme for each of 64 sub-bands. The residual samples in sub-bands are encoded using a multitude of adaptive prediction, scalar/vector quantization and/or Huffman coding to produce the residual codes and extension side information. The packer assembles this data into an extension bit stream.

7.2 DTS Core+96 kHz Extension decoder

On the decoder side core and extension parts of the encoded bit stream are fed to their respective sub-band decoders. The reconstructed core sub-band samples are added to the corresponding residual sub-band samples in lower 32 bands. The reconstructed residual sub-band samples in the upper 32 bands remain unaltered. Passing the resulting extension sub-band samples through the synthesis 64-band QMF filter bank produces the 96 kHz sampled PCM audio. Figure 7.3 shows the block diagram of the core+extension decoder.

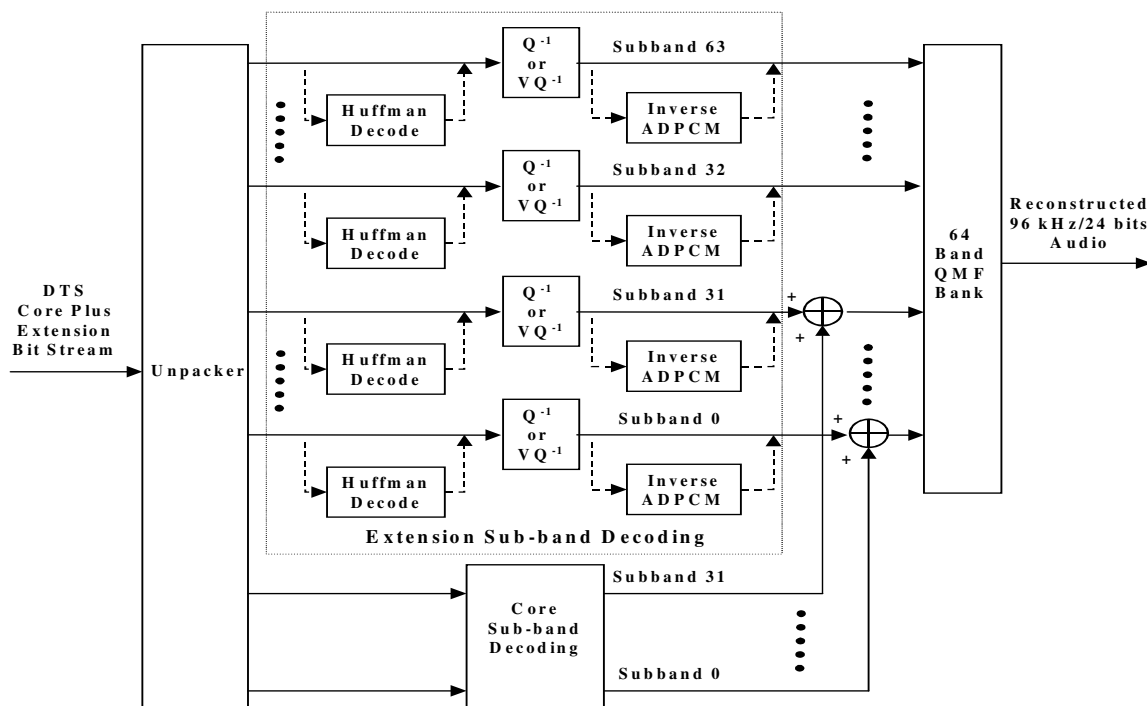


Figure 7.3: The block diagram of DTS Core+Extension decoder

In the case where the encoded bit stream does not contain the extension data, the decoder based on its hardware configuration uses: a) a 32-band QMF with core sub-band samples as inputs to synthesize the 48 kHz sampled PCM audio; b) a 64-band QMF with inputs being core sub-band samples in the lower 32 bands and "zero" samples in the upper 32 bands to synthesize the interpolated PCM audio sampled at 96 kHz.

The existing DTS core decoders when receiving the core+extension bit stream will extract and decode the core data to produce the 48 kHz sampled PCM audio. The decoder ignores the extension data by skipping the extraction until the next DTS synchronization word.

7.3 Synchronization

96 kHz Extension Sync Word SYNC96 V 32 bits

The synchronization word SYNC96 = 0x1D95F262 for the 96 kHz extension data comes after the core audio data. Note that if a channel extension is present the X96k extension data is placed before the XCh extension data in the encoded bit stream. For 16-bit streams the sync word is aligned to 32-bit word boundary. In the case of 14-bit streams SYNC96 is aligned to both 32-bit and 28-bit word boundaries meaning that 28 MSB-s of the SYNC96 appear as 0x07651F26.

To reduce the probability of false synchronization caused by the presence of pseudo sync words, it is imperative to check the distance between the detected sync word and the end of current frame (as indicated by FSIZE). This distance in bytes must match the value of FSIZE96 (see below).

After the decoder synchronization is established a flag `nX96kPresent` is set and the decoder output sampling frequency is selected as

Pseudo Code: `OutSamplingFreq = SFREQ`

`if (nX96kPresent)`

`OutSamplingFreq = 2 × OutSamplingFreq`

Note that `SFREQ` corresponds to a sampling frequency of reconstructed audio in the core decoder.

7.4 X96k frame header

96 kHz Extension Frame Byte Data Size `FSIZE96` **V** **12 bits**

(`FSIZE96 + 1`) is the byte size of 96 kHz extension data plus any other extension data that appears in between `FSIZE96` and the end of current frame. Valid range for `FSIZE96`: 95 to 4 095; Invalid range: 0 to 94.

Revision Number `REVNO` **ACC/NV** **4 bits**

Revision number for the high frequency extension processing algorithm.

Table 7.1: X96k Algorithm Revision Number

REVNO	Frequency Extension Encoder Software Revision Number
0	Reserved
1	Current
2 to 7	Future revision (compatible with the original Rev1.0 specification)
8 to 15	Future revision (incompatible with the original Rev1.0 specification)

NOTE: If the decoder is not compatible with some algorithm revisions (`REVNO > 7`) it must ignore the X96k extension stream and reconstruct the core encoded audio components up to 24/22,05 kHz.

List of Tables

Table 5.1: Frame Type	11
Table 5.2: Deficit Sample Count	12
Table 5.3: CRC Present Flag	12
Table 5.4: Audio channel arrangement	13
Table 5.5: Core audio sampling frequencies	13
Table 5.6: Sub-sampled audio decoding for standard sampling rates	14
Table 5.7: RATE parameter vs. targeted bit-rate.....	14
Table 5.8: Targeted and actual bit-rate for the CD and DVD-Video applications	15
Table 5.9: Status of embedded down mixing coefficients.....	15
Table 5.10: Embedded Dynamic Range Flag.....	15
Table 5.11: Embedded Time Stamp Flag	15
Table 5.12: Auxiliary Data Flag.....	15
Table 5.13: Extension Audio Descriptor Flag	16
Table 5.14: Extended Coding Flag	16
Table 5.15: Audio Sync Word Insertion Flag.....	16
Table 5.16: Flag for LFE channel	16
Table 5.17: Multirate interpolation filter bank switch.....	17
Table 5.18: Encoder software revision	17
Table 5.19: Quantization resolution of source PCM samples	17
Table 5.20: Sum/difference decoding status of front left and right channels	18
Table 5.21: Sum/difference decoding status of left and right surround channels.....	18
Table 5.22: Dialog Normalization Parameter.....	18
Table 7.1: X96k Algorithm Revision Number	23

Annex A (informative): Bibliography

Zoran Fejzo: "DTS Coherent Acoustics; Core and Extensions, Overview of Technology and Description of DTS Stream Frame Headers"

DTS, Inc. (5171 Clareton Drive Agoura Hills, CA 91301): "DTS Decoder Manual Rev2.1 and it's Amendment Rev1.1"

History

Document history		
V1.1.1	August 2002	Publication